

## CLAIMS

What is claimed is:

5

1. A method for determining a plurality of hypothetical matches to a spoken input, comprising the computer-implemented steps of:

detecting subword units in the spoken input to generate a first set of hypothetical matches to the spoken input;

10

detecting words in the spoken input to generate a second set of hypothetical matches to the spoken input; and

combining the first set of hypothetical matches with the second set of hypothetical matches to produce a combined set of hypothetical matches to the spoken input, the combined set having a predefined number of hypothetical matches.

15

2. The method of Claim 1, wherein the step of detecting subword units includes detecting the subword units in the spoken input based on an acoustic model of the subword units and a language model of the subword units;

20

generating pattern comparisons between (i) an input pattern corresponding to the subword units in the spoken input and (ii) a source set of reference patterns based on a pronunciation dictionary, each generated pattern comparison based on the input pattern and one of the reference patterns; and

25

generating the first set of the hypothetical matches by sorting the source set of reference patterns based on a closeness of each reference pattern to correctly matching the input pattern based on an evaluation of each generated pattern comparison, each evaluation determining a word pronunciation distance measure that indicates how close each input pattern is to matching each reference pattern.

30

1001797 "12101  
FOOT" 622001

3. The method of Claim 1, wherein the combined set of hypothetical matches is an ordered list comprising a highest ranking hypothetical match in the second set of hypothetical matches, followed by an ordered set of hypothetical matches based on the first set of hypothetical matches.
- 5
4. The method of Claim 1, wherein the combined set of hypothetical matches is an ordered list based on ranking confidence levels for each hypothetical match.
- 10
5. The method of Claim 1, wherein the subword units include at least one phoneme.
6. The method of Claim 1, wherein the hypothetical matches are words.
- 15
7. A computer system for determining a plurality of hypothetical matches to a spoken input, comprising:
- a subword decoder for detecting subword units in the spoken input to generate a first set of hypothetical matches to the spoken input;
- a word decoder detecting words in the spoken input to generate a second set of hypothetical matches to the spoken input; and
- 20
- a list fusion module for combining the first set of hypothetical matches with the second set of hypothetical matches to produce a combined set of hypothetical matches to the spoken input, the combined set having a predefined number of hypothetical matches.
- 25
8. The computer system of Claim 7, wherein the subword decoder
- detects the subword units in the spoken input based on an acoustic model of the subword units and a language model of the subword units;
- generates pattern comparisons between (i) an input pattern corresponding to the subword units in the spoken input and (ii) a source set of reference
- 30
- patterns based on a pronunciation dictionary, each generated pattern comparison based on the input pattern and one of the reference patterns; and

generates the first set of the hypothetical matches by sorting the source set of reference patterns based on a closeness of each reference pattern to correctly matching the input pattern based on an evaluation of each generated pattern comparison, each evaluation determining a word pronunciation distance measure that indicates how close each input pattern is to matching each reference pattern.

9. The computer system of Claim 7, wherein the combined set of hypothetical matches is an ordered list comprising a highest ranking hypothetical match in the second set of hypothetical matches, followed by an ordered set of hypothetical matches based on the first set of hypothetical matches.
10. The computer system of Claim 7, wherein the combined set of hypothetical matches is an ordered list based on ranking confidence levels for each hypothetical match.
11. The computer system of Claim 7, wherein the subword units include at least one phoneme.
12. The computer system of Claim 7, wherein the hypothetical matches are words.
13. A computer program product comprising:
  - a computer usable medium for determining a plurality of hypothetical matches to a spoken input; and
  - a set of computer program instructions embodied on the computer useable medium, including instructions to:
    - detect subword units in the spoken input to generate a first set of hypothetical matches to the spoken input;
    - detect words in the spoken input to generate a second set of hypothetical matches to the spoken input; and

combine the first set of hypothetical matches with the second set of hypothetical matches to produce a combined set of hypothetical matches to the spoken input, the combined set having a predefined number of hypothetical matches.

5

14. A method for determining a plurality of hypothetical matches to a spoken input by detecting subword units in the spoken input, comprising the computer-implemented steps of:

10 detecting the subword units in the spoken input based on an acoustic model of the subword units and a language model of the subword units;  
generating pattern comparisons between (i) an input pattern corresponding to the subword units in the spoken input and (ii) a source set of reference patterns based on a pronunciation dictionary, each generated pattern comparison based on the input pattern and one of the reference patterns; and  
15 generating a set of the hypothetical matches by sorting the source set of reference patterns based on a closeness of each reference pattern to correctly matching the input pattern based on an evaluation of each generated pattern comparison, each evaluation determining a word pronunciation distance measure that indicates how close each input pattern is to matching each  
20 reference pattern.

15. The method of Claim 14, wherein the pattern comparisons are based on a confusion matrix that stores the likelihood of confusion between pairs of subword units, the likelihood of deleting each subword unit, and the likelihood  
25 of inserting each subword unit.

16. The method of Claim 15, further comprising a step of training the confusion matrix based on an output of an subword decoder, the output produced from an acoustic input of a training data set input to the subword decoder.

30

FOIA b 7 - D

17. The method of Claim 15, further comprising the step of computing the confusion matrix by determining an entry in the confusion matrix for each unique subword unit that is in the set of reference patterns.
- 5 18. The method of Claim 14, wherein the step of detecting subword units is performed in a client computer, and the steps of generating pattern comparisons and generating the set of hypothetical matches are performed in a server computer.
- 10 19. The method of Claim 14, wherein the step of generating the set of hypothetical matches includes:
- determining pairs of subword units by pairing an input subword unit from the input pattern with a reference subword unit from the reference pattern; and
- 15 providing the word pronunciation distance measure by calculating a distance metric for each pair of subword units, the distance metric defined as follows:

$$\begin{aligned}
 S(p_0, d_0) &= 0 \\
 S(p_i, d_j) &= \min \left\{ \begin{aligned} &S(p_{i-1}, d_{j-1}) + C_{subs}(p_i, d_j) \\ &S(p_{i-1}, d_j) + C_{del}(p_i) \\ &S(p_i, d_{j-1}) + C_{ins}(d_j) \end{aligned} \right\} \\
 S(P, D) &= S(p_n, d_m) + LP(p_n, d_m)
 \end{aligned}$$

20

wherein:

$S(P, D)$  is a distance between word P and D; P is a given input pattern, and D, a given reference pattern;

25  $S(p_i, d_j)$  is a score of the given input pattern matching a given subword unit  $p_i$  of P, and a given subword unit  $d_j$  of D;

$C_{subs}(p_i, d_j)$  is a cost of substituting the given subword unit  $p_i$  of P with the given subword unit  $d_j$  of D;

10017797-121201

$C_{del}(p_i)$  is a cost of deleting the given subword unit  $p_i$  of  $P$ ;

$C_{ins}(d_j)$  is a cost of inserting the given subword unit  $d_j$  of  $D$ ;

$LP(p_n, d_m)$  is a length penalty of the given input pattern  $p_n$  matching the given reference pattern  $d_m$ ,  $n$  is the length of  $P$ , and  $m$  is the length of  $D$ ;

5  $S(p_{i-1}, d_{j-1})$  has a value of zero (0) if  $p_{i-1}, d_{j-1}$  is undefined;

$S(p_{i-1}, d_j)$  has the value of zero (0) if  $p_{i-1}, d_j$  is undefined;

$S(p_i, d_{j-1})$  has the value of zero (0) if  $p_i, d_{j-1}$  is undefined; and

the distance metric for each pair of subword units is calculated in a sequence such that  $S(p_{i-1}, d_{j-1})$ ,  $S(p_{i-1}, d_j)$ , and  $S(p_i, d_{j-1})$  are determined previously to  
10 determining  $S(p_i, d_j)$ .

20. The method of Claim 14, wherein the subword units include at least one phoneme.

15 21. An computer system for determining a plurality of hypothetical matches to a spoken input by detecting subword units in the spoken input, comprising:

a subword decoder for detecting the subword units in the spoken input based on an acoustic model of the subword units and a language model of the subword units; and

20 a subword detection vocabulary look up module for generating pattern comparisons between (i) an input pattern corresponding to the subword units in the spoken input and (ii) a source set of reference patterns based on a pronunciation dictionary, each generated pattern comparison based on the input pattern and one of the reference patterns;

25 the subword detection vocabulary look up module generating a set of the hypothetical matches by sorting the source set of reference patterns based on a closeness of each reference pattern to correctly matching the input pattern based on an evaluation of each generated pattern comparison, each evaluation determining a word pronunciation distance measure that indicates how close  
30 each input pattern is to matching each reference pattern.

FOIA b 7 - D

22. The computer system of Claim 21, wherein the pattern comparisons are based on a confusion matrix that stores the likelihood of confusion between pairs of subword units, the likelihood of deleting each subword unit, and the likelihood of inserting each subword unit.
23. The computer system of Claim 22, wherein the confusion matrix is trained on an output of the subword decoder, the output produced from an acoustic input of a training data set to the subword decoder.
24. The computer system of Claim 22, wherein the confusion matrix is based on determining an entry in the confusion matrix for each unique subword unit that is in the set of reference patterns.
25. The computer system of Claim 21, wherein the subword decoder is part of a client computer, and the subword detection vocabulary look up module is part of a server computer.
26. The computer system of Claim 21, wherein the subword detection vocabulary look up module  
determines pairs of subword units by pairing an input subword unit from the input pattern with a reference subword unit from the reference pattern; and  
provides the word pronunciation distance measure by calculating a distance metric for each pair of subword units, the distance metric defined as follows:

$$S(p_0, d_0) = 0$$

$$S(p_i, d_j) = \min \begin{cases} S(p_{i-1}, d_{j-1}) + C_{subs}(p_i, d_j) \\ S(p_{i-1}, d_j) + C_{del}(p_i) \\ S(p_i, d_{j-1}) + C_{ins}(d_j) \end{cases}$$

$$S(P, D) = S(p_n, d_m) + LP(p_n, d_m)$$

wherein:

$S(P,D)$  is a distance between word  $P$  and  $D$ ;  $P$  is a given input pattern, and  $D$ , a given reference pattern;

5  $S(p_i, d_j)$  is a score of the given input pattern matching a given subword unit  $p_i$  of  $P$ , and a given subword unit  $d_j$  of  $D$ ;

$C_{subs}(p_i, d_j)$  is a cost of substituting the given subword unit  $p_i$  of  $P$  with the given subword unit  $d_j$  of  $D$ ;

$C_{del}(p_i)$  is a cost of deleting the given subword unit  $p_i$  of  $P$ ;

$C_{ins}(d_j)$  is a cost of inserting the given subword unit  $d_j$  of  $D$ ;

10  $LP(p_n, d_m)$  is a length penalty of the given input pattern  $p_n$  matching the given reference pattern  $d_m$ ,  $n$  is the length of  $P$ , and  $m$  is the length of  $D$ ;

$S(p_{i-1}, d_{j-1})$  has a value of zero (0) if  $p_{i-1}, d_{j-1}$  is undefined;

$S(p_{i-1}, d_j)$  has the value of zero (0) if  $p_{i-1}, d_j$  is undefined;

$S(p_i, d_{j-1})$  has the value of zero (0) if  $(p_i, d_{j-1})$  is undefined; and

15 the distance metric for each pair of subword units is calculated in a sequence such that  $S(p_{i-1}, d_{j-1})$ ,  $S(p_{i-1}, d_j)$ , and  $S(p_i, d_{j-1})$  are determined previously to determining  $S(p_i, d_j)$ .

20 27. The computer system of Claim 21, wherein the subword units include at least one phoneme.

28. A computer program product comprising:  
a computer usable medium for determining a plurality of hypothetical matches to a spoken input by detecting subwords in the spoken input; and  
25 a set of computer program instructions embodied on the computer useable medium, including instructions to:

30 detect the subword units in the spoken input based on an acoustic model of the subword units and a language model of the subword units;  
generate pattern comparisons between (i) an input pattern corresponding to the subword units in the spoken input and (ii) a source set of reference patterns based on a pronunciation dictionary, each



generated pattern comparison based on the input pattern and one of the reference patterns; and

5 generate a set of the hypothetical matches by sorting the source set of reference patterns based on a closeness of each reference pattern to correctly matching the input pattern based on an evaluation of each generated pattern comparison, each evaluation determining a word pronunciation distance measure that indicates how close each input pattern is to matching each reference pattern.

10

1001797-121001